COMinDS 2019 Workshop - Titles and Abstracts

Francis Bach

Title: **Machine Learning over Networks**

Abstract: The success of machine learning models is in part due to their capacity to train on large amounts of data. Distributed systems are the common way to process more data than one computer can store, but they can also be used to increase the pace at which models are trained by splitting the work among many computing nodes. In this talk, I will study the corresponding problem of minimizing a sum of functions which are respectively accessible by separate nodes in a network. New centralized and decentralized algorithms will be presented, together with their convergence guarantees in deterministic and stochastic convex settings, leading to optimal algorithms for this particular class of distributed optimization problems.

Ulrich Bauer

Title: **Persistent homology: from theory to computation**

Abstract: In this talk, I will give an overview on some theoretical and computational results in applied topology. I will focus on three aspects of persistent homology as a topological descriptor: its use for the inference and simplification of topological features, its stability with respect to perturbations of the data, and efficient methods for its computation on a large scale. These questions will be motivated and illustrated by concrete examples and problems, such as - reconstruction of a shape and its topological properties from a point cloud, - denoising of isosurfaces for the visualization of medical images, and - faithful simplification of contours lines of a terrain.

Peter Benner

Title: **Data-enhanced Reduced-order Modeling for Dynamical Systems**

Abstract: The design of surrogate models for faster simulation, optimization and control of dynamical systems has become an active research area within CSE during the last two decades. Model order reduction (MOR) is the prevailing technique to compute a reduced-order model as a surrogate for a high-dimensional (full-order) model, using often some form of (Petrov-)Galerkin projection. This requires the availability of a model description, usually in matrix format, for the full-order model. Recently, inspired by the successes of data science and artificial intelligence in other areas, more and more focus is given to deriving surrogate models from data, using so-called non-intrusive methods that do not require knowledge of the full-order model, but are based on available time series data or on an oracle producing suitable output data, given input data. We will provide a survey on recent activities in this direction.

Alexandra Carpentier

Title: **Adaptive inference and its relations to sequential decision making**

Abstract: (based on joint works with Andrea Locatelli, Matthias Loeffler, Olga Klopp and Richard Nickl)

Adaptive inference - namely adaptive estimation and adaptive confidence statements - is particularly important in high or infinite dimensional models in statistics. Indeed whenever the dimension becomes high or infinite, it is important to adapt to the underlying structure of the problem. While adaptive estimation is often possible, it is often the case that adaptive and honest confidence sets do not exist. This is known as the adaptive inference paradox. And this has consequences in sequential decision making. In this talk, I will present some classical results of adaptive inference and discuss how they impact sequential decision making.

Hanno Gottschalk

Title: **Meta Classification to detect Errors in Semantic Segmentation**

Abstract: The application of machine learning technology in safety relevant environments poses several challenges regarding their reliability. This especially applies to modern deep learning architectures, which pretty much resemble a black box. In this talk, measures of uncertainty from basic classification algorithms to current fully convolutional networks for semantic segmentation are discussed. We present known and new uncertainty measures and use them in meta classification: An algorithmic decision, if a prediction by the AI is correct of not, without access to the ground truth data. For the case of semantic segmentation, we use this for the instance based detection of false positives. Techniques for the detection of false negatives are discussed as well. Both sets of techniques also prove useful for active learning and quality control in data annotation.

Matthias Hein

Title: **Towards robust and safe learning**

Abstract: Machine Learning and in particular, deep learning, needs to satisfy high requirements on reliability and robustness when applied in safety-critical applications. However, at the moment neural networks are non-robust, as small adversarial changes of the input change the decision of a classifier, and non-reliable, as neural networks make high-confidence predictions far away from the training data. I will present our work towards provable robustness of neural networks and how to overcome the problem that neural networks don't know, when they don't know.

Arnulf Jentzen

Title: **Overcoming the curse of dimensionality: from nonlinear Monte Carlo to deep artificial neural networks**

Abstract: Partial differential equations (PDEs) are among the most universal tools used in modelling problems in nature and man-made complex systems. For example, stochastic PDEs are a fundamental ingredient in models for nonlinear filtering problems in chemical engineering and weather forecasting, deterministic Schroedinger PDEs describe the wave function in a quantum physical system, deterministic Hamiltonian-Jacobi-Bellman PDEs are employed in operations research to describe optimal control problems where companys aim to minimise their costs, and deterministic Black-Scholes-type PDEs are highly employed in portfolio optimization models as well as in state-of-the-art pricing and hedging models for financial derivatives. The PDEs appearing in such models are often high-dimensional as the number of dimensions, roughly speaking, corresponds to the number of all involved interacting substances, particles, resources, agents, or assets in the model. For instance, in the case of the above mentioned financial engineering models the dimensionality of the PDE often corresponds to the number of financial assets in the involved hedging portfolio. Such PDEs can typically not be solved explicitly and it is one of the most challenging tasks in applied mathematics to develop approximation algorithms which are able to approximatively compute solutions of high-dimensional PDEs. Nearly all approximation algorithms for PDEs in the literature suffer from the so-called "curse of dimensionality" in the sense that the number of required computational operations of the approximation algorithm to achieve a given approximation accuracy grows exponentially in the dimension of the considered PDE. With such algorithms it is impossible to approximatively compute solutions of high-dimensional PDEs even when the fastest currently available computers are used. In the case of linear parabolic PDEs and approximations at a fixed space-time point, the curse of dimensionality can be overcome by means of Monte Carlo approximation algorithms and the Feynman-Kac formula. In this talk we prove that suitable deep artificial neural network approximations do indeed overcome the curse of dimensionality in the case of a general class of semilinear parabolic PDEs and we thereby prove, for the first time, that a general semilinear parabolic PDE with a nonlinearity depending on the PDE solution can be solved approximatively without the curse of dimensionality.

Klaus-Robert Müller

Title: **Explainable AI and Applications**

Abstract: We first provide a short introduction into techniques for explainable AI for deep learning. Subsequently, applications from the sciences are discussed.

Frank Noé

Title: **Machine Learning for fundamental physics problems**

Abstract: Machine learning (ML) and AI have made impressive strides in the past few years. We are developing ML methods to address fundamental problems in physics. Such learning structures differ from generic ML methods as physical theory defines constraints, symmetries and conservation laws that a reasonable ML model for a physical phenomenon must obey. Here we present two examples of physical problems where developing ML structures with the right physics built in can make sizeable progress: the sampling problem in statistical physics, and finding quantum-mechanical ground states in quantum mechanics.

References:
`https://science.sciencemag.org/content/365/6457/eaaw1147`,
`https://arxiv.org/abs/1909.08423`

Philipp Petersen

Title: **Deep neural networks, partial differential equations, overparametrisation, and convergence**

Abstract: Deep neural networks are nowadays used to solve various problems of applied mathematics. One instance that is particularly interesting from a mathematical point of view is their application to the solution of PDEs. The dominant framework here is to apply a collocation method that amounts to minimising a discretised energy. We will discuss a couple of issues with this approach that prevent a theory of convergence. Finally, we will identify an alternative class of differential equations that can be solved by a deep neural network-based techniques and do not admit these problems.

Holger Rauhut

Title: **Learning deep linear neural networks: Riemannian gradient flows and convergence to global minimizers**

Abstract: Abstract: We study the convergence of gradient flows related to learning deep linear neural networks from data. In the linear case, the composition of the network layers amounts to simply multiplying the weight matrices of all layers together, resulting in an overparameterized problem. We show that the gradient flow with respect to these factors can be re-interpreted as a Riemannian gradient flow on the manifold of rank-r matrices endowed with a suitable Riemannian metric. We show that the flow always converges to a critical point of the underlying functional. Moreover, in the special case of an autoencoder, we show that the flow converges to a global minimum for almost all initializations.

Sebastian Reich

Title: **Erwin Schroedinger and Data Science**

Abstract: In 1931, Erwin Schroedinger posed a certain boundary value problem in the space of probability measures which has recently found connections and applications to a wide range of problems arising from data science including data assimilation, clustering of data, and learning. In my talk I will summarise this Schroedinger perspective on data assimilation and its link to the classical Kalman filter for stochastic processes; a more detailed exposition of which can be found in my 2019 Acta Numerica paper (arXiv:1807.08351).

Sebastian Sager

Title: **Expert-Enhanced Machine Learning for Cardiac Arrhythmia Classification**

Abstract: Machine learning (ML) methodology has been successfully applied to many classification problems in medicine and beyond. Whereas the accuracy is often astonishing, the interpretability of the results has become an ubiquitous issue. In order to overcome this important but unsolved challenge, we propose to first reduce the complexity of the data, and then to combine the interpretability of expert systems with the deductive power of data driven ML. As a showcase we considered the arguably most difficult classification case of cardiac arrhythmias. Here the largest database with the gold standard (intracardiac measurements after invasive procedures) only contains 380 samples, yielding an additional challenge to ML. Still, our approach achieved an accuracy of 82.84%. The main advantage however is the interpretability of the classification results. Our features give insight into a possibly occurring multi-level atrioventricular blocking mechanism, which might improve treatment decisions and is thus an important step in the realization of personalized medicine. The idea to use mathematical modeling and optimization to generate new and clinically interpretable features for ML can be transferred to other cases of clinical decision support.

Gabriele Steidl

Title: **Curve based approximation of measures on manifolds by discrepancy minimization**

Abstract: The approximation of probability measures on compact metric spaces and in particular on Riemannian manifolds by atomic or empirical ones is a classical task in approximation and complexity theory with a wide range of applications. Instead of point measures we are concerned with the approximation by measures supported on Lipschitz curves. Special attention is paid to push-forward measures of Lebesgue measures on the interval by such curves. Using the discrepancy as distance between measures we provide approximation rates in terms of Lipschitz constants of curves. For general probability measures on compact metric spaces with few additional properties we prove approximation rates based on those for atomic measure approximation and cost estimates of the traveling salesman problem. These approximation rates can be improved for absolutely continuous measures on compact d-dimensional Riemannian manifolds having densities in certain Sobolev spaces. Here our approach relies on the construction of suitable quadrature measures.

Having established the theoretical convergence rates, we are interested in the numerical minimization of the discrepancy between a given probability measure and the set of push-forwards of Lebesgue measures on the interval by Lipschitz curves. We present numerical examples for measures on the d-dimensional torus, the 2-sphere, the rotation group on R^3 and the Grassmannian of all two-dimensional linear subspaces of ^{4}. As algorithm of choice we choose a CG method on these manifolds which directional updates incorporate second-oder information. For efficiently computing the gradients and the Hessian within the algorithm we approximate the given measures by finite sums of eigenfunctions of the Laplace-Beltrami operator and use fast Fourier transform techniques on these manifolds.

This is joint work with Martin Ehler (University of Vienna), Manuel Gräf (Austrian Academy of Sciences) and Sebastian Neumayer (TU Kaiserslautern).

Bernd Sturmfels

Title: **Statistical Models with Rational Maximum Likelihood Estimator**

Abstract: A statistical model for discrete data is a subset of a probability simplex. Its maximum likelihood estimator (MLE) is a retraction from the simplex to the model. We characterize all models for which the MLE is a rational function. This rests on real algebraic geometry results, mostly due to Huh and Kapranov. Our discussion centers around models used in practise, like Bayesian networks, graphical models, and staged trees. This is joint work with Eliana Duarte and Orlando Marigliano.